Virology **304,** 311–329 (2002)
doi:10.1006/viro.2002.1691

The Sequential Introduction of HIV-1 Subtype B and CRF01_AE in Singapore by Sexual Transmission: Accelerated V3 Region Evolution in a Subpopulation of Asian CRF01 Viruses

Marcia L. Kalish,*[,1] Bette T. Korber,† Satish Pillai,† Kenneth E. Robbins,* Yee Sin Leo,‡ Ae Saekhou,* Iris Verghese,‡ Phillip Gerrish,*[,2] Chee Leok Goh,§ Davis Lupo,* Boon Huan Tan,§[,3] Teresa M. Brown,* and Roy Chan§

*Centers for Disease Control and Prevention, Atlanta, Georgia 30333; †Los Alamos National Laboratory, Los Alamos, New Mexico 87545; ‡Communicable Disease Centre, Singapore 308205; and §National Skin Centre, Singapore

Received April 16, 2002; returned to author for revision May 28, 2002; accepted July 26, 2002

The rapid spread of the human immunodeficiency virus type 1 (HIV-1) circulating recombinant form (CRF) 01_AE throughout Asia demonstrates the dynamic nature of emerging epidemics. To further characterize the dissemination of these strains regionally, we sequenced 58 strains from Singapore and found that subtype B and CRF01 were introduced separately, by homosexual and heterosexual transmission, respectively. Protein similarity scores of the Singapore CRF01, as well as all Asian strains, demonstrated a complex distribution of scores in the V3 loop—some strains had very similar V3 loop sequences, while others were highly divergent. Furthermore, we found a strong correlation between the loss of a V3 glycosylation site and the divergent strains. This suggests that loss of this glycosylation site may make the V3 loop more susceptible to immune surveillance. The identification of a rapidly evolving population of CRF01_AE variants should be considered when designing new candidate vaccines and when evaluating breakthrough strains from current vaccine trials.

Key Words: HIV-1; Singapore; CRF01_AE; V3 loop; envelope sequences; diversity; glycosylation sites; risk group; evolution; Asia.

INTRODUCTION

The greatest diversity of human immunodeficiency viruses has been identified in Africa (*17, 20, 29, 47, 56*), where the introduction of these retroviruses into humans was due to zoonotic infections (*11, 22, 24, 60, 61*). After human immunodeficiency virus type 1 (HIV-1) variants radiated out from Africa, numerous regional epidemics occurred worldwide. Often the variants that dominated these regional expansions had a low prevalence in Africa. For example, subtype B viruses, which are infrequently found in Africa (*40*), were the first identified HIV-1 strains (*1, 18*), detected after their dissemination into the Americas and Europe (*28, 55*). Subtype F1 viruses (*80*) were first recognized after seeding an epidemic of pediatric infections in Romania (*16*). One of the most intensively studied of these regional epidemics, and the subject of this investigation, was the introduction and rapid spread throughout Asia of subtype E viruses, now called CRF01 (circulating recombinant form (CRF) 01_AE or CRF01), with CM240 being a prototype strain CM240 (*65, 66*). CRF01 isolates from 11 different countries in Asia are currently available in the HIV sequence database (*40*). The first reported strains of CRF01 were from Thailand (*54, 58, 59*), and these strains were later found to be related to a more diverse set of CRF01 viruses in the Central African Republic (*56*). This variant is thought to be a mosaic since its *gag*, *pol*, and most of its accessory genes cluster with subtype A, but because most of the external portion of *gp*160 and part of its LTR are very distinctive from those of other subtype A viruses, it was given the designation CRF (*6, 21, 66, 67*).

The epidemic in Southeast Asia was first recognized in Thailand in 1988, when the prevalence of HIV (later identified as subtype B′) among injecting drug users (IDUs) in Bangkok increased from <1 to 43% in 1 year (*10*). By the following year, a similar explosive rise in infections among female commercial sex workers (CSWs) was noted, a rise that heralded the start of the heterosexual HIV CRF01 epidemic (*5, 50, 71, 72, 82*). The sequences from the early CRF01 epidemic in Thailand were very conserved, particularly relative to the CRF01 sequences from the Central African Republic, suggesting that a founder virus from the older African epidemic seeded the CRF01 epidemic in Asia (*55, 56, 77*). What originally appeared to be sequential waves of infections beginning among IDUs and spreading sexually was later recognized as separate, near-simultaneous introductions of subtype B′ viruses through parenteral transmis-

[1] To whom correspondence and reprints requests should be addressed at Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop G19, Atlanta, Georgia 30333. Fax: (404) 639-3254. E-mail: mkalish@cdc.gov.

[2] Present address: Theoretical Biology & Biophysics, MailStop K710, Los, Alamos National Laboratory, Los Alamos, New Mexico 87545.

[3] Present address: Defense Medical Research Institute, Singapore.

0042-6822/02 $35.00

sions and of CRF01 viruses through sexual transmissions (*58, 59*). A later study of the molecular epidemiology of HIV among Bangkok IDUs highlighted the dynamic nature of the dispersal of HIV strains within a population. By 1994, the proportion of CRF01 infections among IDUs infected with HIV less than 2 years had increased to 44% (*31*), and from 1995–1998 the proportion in new CRF01 IDU infections had increased to 80% (*78, 81*).

To further study the dynamics of the emerging Asian epidemic, as well as to track patterns in the evolution of the viruses, we investigated the molecular epidemiology and evolutionary relationships of HIV in Singapore. We have documented that subtype B was transmitted primarily by homosexual transmission and CRF01-AE by heterosexual transmission. Further, we show that a subpopulation of CRF01 viruses are evolving at an accelerated rate in the V3 region and that this rapid adaptation is strongly correlated with the loss of a glycosylation site within the V3 loop.

RESULTS

Patient demographics and clinical histories

Table 1 summarizes the demographic and clinical data from the questionnaires filled out at the time the patients underwent phlebotomy. Of the 58 samples that were sequenced and analyzed, 26 samples were from heterosexuals, 19 samples were from homosexuals, 10 samples were from bisexuals, 2 samples were from patients with both IDU and heterosexual risk factors, and 1 sample was from a patient whose only risk for infection was an organ transplant. Most of the patients were Chinese (86%) and born in Singapore (79%). All but three (95%) of the patients infected sexually were male, and 85% reported having sex in other countries, primarily Thailand, Malaysia, and Indonesia. Large numbers of sex partners were reported regardless of risk group: 14 (25%) of the 55 persons infected sexually reported having more than 50 partners; 18 (33%) reported 20–50 contacts; 9 (16%) reported 10–20 contacts; 7 (13%) reported 5–10 contacts; and 7 (13%) reported fewer than 5 contacts. All three of the women infected sexually reported having fewer than five sex partners, and only one reported having sex outside of Singapore (in Thailand). For most of the respondents, the date of infection was unknown, so only the date of first diagnosis of HIV infection is listed in Table 1. The earliest documented infection among participants in this study was in a homosexual male, known to have been infected since 1988. The average age of all patients was 39.5 years, and their mean CD4 count was 212.5. Patients infected with CRF01 had a mean age of 43 years (range 24–61) and a mean CD4 count of 237 (range 4–769); those with subtype B viruses had a mean age of 35.5 years (range 27–48) and a CD4 count of 189 (8–603). Using a standard analysis of variance, we found the average age of patients infected with CRF01 viruses was significantly higher ($P = 0.0025$) than that of patients infected with subtype B viruses, but we found no significant difference in their CD4 counts ($P = 0.31$).

Through phylogenetic analysis of the 58 env sequences, we found that strains fell into three different group M subtypes or CRFs: B, CRF01, or C. Twenty-nine of the sequences represented CRF01 infections, 28 were subtype B, and 1 was subtype C. Only 3 (11%) of the 28 subtype B infections were due to the genetically distinct subtype B′ variant (*30*). The one subtype C infection was in a woman whose only risk for HIV infection was receiving an organ transplant in India, where subtype C infections dominate the HIV epidemic (*46*). Eighteen (95%) of the 19 persons identifying themselves as having homosexually acquired infections were infected with subtype B viruses, but only 1 of these 18 with subtype B infections had the B′ variant; 23 (88%) of 26 heterosexually transmitted infections were due to CRF01 strains (Fig. 1). Of the 10 bisexuals, half (50%) were infected with CRF01 and half were infected with subtype B viruses. The two persons identifying themselves as IDUs were infected with subtype B′ viruses, typical of the strains reported to be predominant among Bangkok IDUs early in the Thai epidemic. One B′ variant was found in a homosexual male.

Phylogenetic analysis of the C2V3C3 region of gp120

To evaluate the evolutionary relationships of the Singapore HIV-1 strains, we performed phylogenetic analysis on the C2V3C3 region of env and the p17 coding region of gag for all strains. The results of a neighbor-joining tree with bootstrapping for the C2V3C3 region is shown in Fig. 2.

By comparing branch lengths, we were able to estimate relative amounts of evolutionary change among different lineages; we conducted bootstrap resampling of 2000 replicate trees to determine the robustness of the branching order and observed groupings. A bootstrapped neighbor-joining tree was constructed using the subtype B reference strains MN and HXB2; the 1994 Thailand (TH) subtype B′ strains TH101, TH113, TH131, TH132, TH133, TH139, TH149, and TH162 (*31*), the Central African Republic (CAR) CRF01 strains CAR17, CAR39, and CAR4071 (*56*); the Thailand 1994 CRF01 strains TH89, TH100, and TH126; CRF01 reference strains CF402, CM240, CARMBA; the subtype C reference strains BW05, ZAM18, ETH2220, and IN21068; and the subtype F reference strain BR020 as an outgroup. Both subtype B and CRF01 C2V3C3 sequences are represented in two separate star phylogenies. The branch lengths in each cluster represent estimates of the extent of nucleotide change that has occurred since the viruses most recently shared a common ancestor. The branches of the subtype B sequences are generally longer than those of

TABLE 1

Demographic, Risk Factor, and Clinical Data

ID	Clade	Risk group	CD4	Year diagnosed	Sex	Age	Race	Sex overseas	No. of sex partners	Country of birth
6218	B	bi	10	Unk	M	34	Ch	TH, I	10–20	SG
6219	E	he	343	1996	M	61	Ch	TH	20–50	SG
6221	E	he	202	1994	F	33	OT	TH	<5	OT
6265	B′	IDU, he	603	1992	M	38	MA	MA	5–10	SG
6268	E	he	20	1995	M	57	Ch	TH, MA, IN, OT	10–20	SG
6269	B	ho	8	1995	M	37	Ch	None	<5	SG
6270	B	ho	91	1995	M	31	Ch	TH, I	>50	SG
6271	B	bi	246	1994	M	36	Ch	AU	10–20	SG
6273	B	he	509	1993	M	41	In	None	5–10	SG
6274	B	ho	71	1994	M	31	Ch	None	5–10	SG
6276	E	bi	4	1995	M	40	Ch	MA, I	>50	SG
6278	B	ho	17	1992	M	40	Ch	None	<5	SG
6280	B	bi	161	1994	M	29	Ch	MA	10–20	MA
6283	B	ho	121	1994	M	44	Ch	EU, AU, OT	>50	SG
6286	C	Organ	138	1995	F	38	Ch	None	Unk	SG
6287	B	bi	110	1995	M	27	In	None	10–20	SG
6290	B′	ho	325	1992	M	35	Ch	None	20–50	SG
6294	B	ho	170	1993	M	37	Ch	None	20–50	Unk
6295	B	he	22	1994	M	37	Ch	TH, I, OT	>50	SG
6296	B	bi	70	1994	M	48	Ch	EU	10–20	SG
6302	B	ho	411	1994	M	31	Ch	U.S.	<5	SG
6303	B	ho	123	1990	M	42	In	EU, AU, OT	10–20	SG
6309	E	he	304	1993	M	61	Ch	TH, MA, I, OT	>50	OT
6315	E	he	94	1996	M	43	In	TH, OT	5–10	Unk
6316	E	he	138	1993	M	33	Ch	TH, MA, OT	20–50	SG
6318	E	he	18	1994	M	34	Ch	TH, MA	20–50	SG
6320	E	bi	117	1995	M	31	Ch	MA, I	>50	SG
6323	E	bi	396	1992	M	41	Ch	None	20–50	SG
6324	E	bi	Unk	1996	M	24	Ch	TH, MA	>50	SG
6325	E	he	243	1995	M	60	Ch	TH, OT	>50	SG
6326	E	he	224	1996	F	49	Ch	None	<5	SG
6327	B	ho	43	1989	M	35	Ch	TH	20–50	SG
6329	B	ho	45	1991	M	33	Ch	TH-I-U.S.-EU-AU	20–50	SG
6330	E	he	77	1991	M	37	Ch	TH	20–50	SG
6332	E	he	769	1995	M	34	Ch	TH	>50	SG
6334	E	bi	346	1996	M	38	Ch	TH	20–50	SG
6335	E	ho	685	1994	M	42	Ch	MA	>50	SG
6336	E	he	16	1995	M	29	Ch	I	20–50	SG
6337	B′	IDU, he	256	1996	M	28	OT	OT	20–50	OT
6339	E	he	Unk	Unk	M	50	Ch	TH	>50	SG
6341	E	he	382	Unk	M	35	Ch	TH, MA	20–50	SG
6346	E	he	355	1995	M	31	Ch	TH, MA	20–50	SG
6347	B	ho	322	1992	M	31	Ch	MA, OT	20–50	SG
6348	B	ho	114	1988	M	39	Ch	EU, U.S.	20–50	SG
6349	E	he	Unk	Unk	M	50	Ch	TH, MA	10–20	MA
6350	E	he	18	1995	M	45	Ch	TH	5–10	SG
6351	B	ho	314	1996	M	33	Ch	EU	Unk	SG
6352	B	ho	343	1996	M	41	Ch	IN, EU, AU	20–50	SG
6353	E	he	212	1995	M	65	Ch	TH, MA, I, AU	20–50	OT
6354	B	ho	Unk	Unk	M		Ch	OT	>50	SG
6355	B	he	42	1991	M	43	Ch	TH, MA	10–20	SG
6357	E	he	8	1994	M	46	Ch	TH, I, MA	>50	OT
6358	B	ho	464	1995	M	28	Ch	TH	5–10	SG
6359	B	ho	339	1996	M	29	Ch	None	5–10	SG
6360	E	he	544	1996	F	26	OT	TH	<5	OT
6362	E	he	594	1995	M	44	Ch	TH, I, MA	>50	OT
6363	E	he	27	1996	F	62	Ch	None	<5	SG
6364	E	he	34	1996	M	56	Ch	TH, I	>50	MA

Note. bi = bisexual; he = heterosexual; ho = homosexual; IDU = injecting drug user; Organ = organ transplant recipient; Ch = Chinese; Ma = Malaysian; In = Indian; OT = other; Unk = unknown; SG = Singapore; TH = Thailand; MA = Malaysia; IN = India; I = Indonesia; EU = Europe; AU = Australia; U.S. = United States.

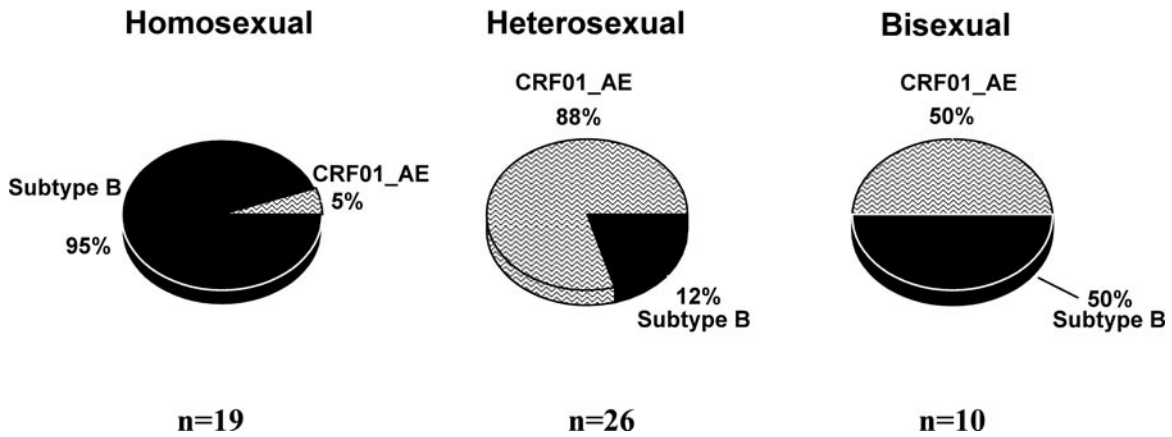

FIG. 1. Subtype distribution in Singapore by risk group. Each pie chart represents the percentage of subtype B and CRF01_AE strains by homosexual, heterosexual, and bisexual risk group.

the Singapore CRF01 sequences, indicating that the subtype Bs are more divergent from each other than are the CRF01 sequences. The longer branches could reflect an earlier introduction or multiple introductions of the subtype B viruses, and these hypotheses are not mutually exclusive. Within subtype B, the B′ viruses form a phylogenetically distinct subcluster (*30*). Through bootstrap reconstruction of the branching order, we demonstrated that all the subtype B′ viruses grouped together in 1680 (84%) of 2000 replicate trees as a subcluster within the general subtype B clade (Fig. 2). The formation of this monophyletic cluster within the B clade reflects the epidemiologic relatedness of these viruses to the Asia founder virus and the recent expansion of these strains. The 1996 Singapore B′ strains have longer branches than the 1994 Thailand B′ strains, reflecting the continued evolution of these viruses.

Within the CRF01 clade, the longer branches that we observed in a subpopulation of Singapore CRF01 viruses (such as SG6316, SC6320, SG6324, SG6325, and SG6330) indicates that these sequences are more divergent than the other 1996 Singapore CRF01 strains. The degree of variation in these viruses is similar to that seen in CRF01 viruses from the Central African Republic, suggesting that there may be different subpopulations of CRF01 viruses within our study population. As will be discussed below, we observed this divergence only in the V3 loop and not in p17. A similar pattern of divergence, in a subset of V3 sequences, was noted previously in CRF01 sequences derived from AIDS patients in northern Thailand (*83*).

Currently, infection with CRF01 viruses has been reported throughout Asia (*40*), including cases in Japan, Myanmar (Burma), Indonesia, Laos, Taiwan, Hong Kong, Vietnam, Cambodia, Malaysia, and China, as well as in Thailand. Some regions of Asia have a high prevalence. When we constructed a C2V3C3 region neighbor-joining tree based on 288 CRF01 sequences extracted from the Los Alamos database, including our sequences along with samples from studies conducted throughout Asia and the Central African Republic, we found no particular clustering of viral sequences associated with the Singapore strains. The Asian CRF01 sequences were intermingled in the tree, although the African and Asian sequences formed distinct clades (Fig. 3). Unlike the available Chinese sequences which form a distinctive subcluster, the Singapore sequences do not appear to be from a single founder virus. Instead, they seem to be part of a more fluid pattern of forms circulating throughout Asia. This is consistent with the epidemiology of HIV being introduced into Singapore from sexual exposures in other Asian countries where CRF01 strains predominate.

As discussed previously, the single subtype C sequence was more closely related to an India subtype C strain than to related viruses from Zambia, Ethiopia, and Botswana. This is consistent with the reports that the only risk factor for infection in the patient, a child, was receiving an organ transplant in India, where subtype C viruses predominate.

Phylogenetic analysis of the p17 coding region of the gag gene

We conducted phylogenetic analysis of p17 sequences (Fig. 4) to compare the tree topology or branching order with that of the C2V3C3 region. Reference strains used in the neighbor-joining tree with bootstrapping were subtype B strains MN and HXB2; 1991 subtype B′ strains TH101, TH132, TH162; and subtype A strains U455 and Q2317. Consistent with reports on the recombinant nature of CRF01 viruses (*6, 21*), p17 sequences that were identified as subtype E in the C2V3C3 region clustered with subtype A sequences (confirming their A/E recombinant form). When compared with other subtype A p17 sequences from Africa, the Singapore sequences formed a distinct phylogenetic subcluster in 1880 (94%) of 2000 bootstrap replicates, reflecting their

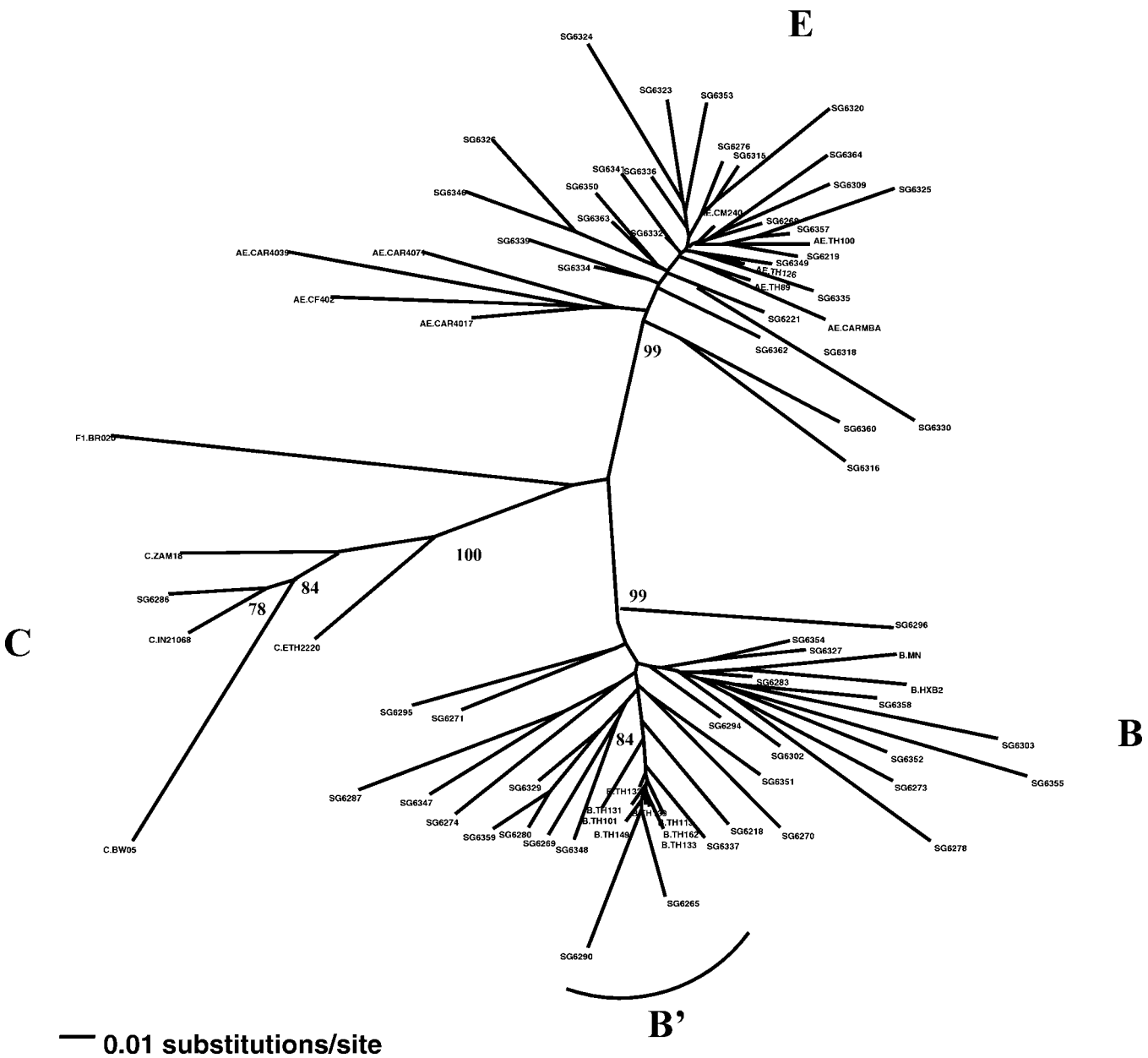

FIG. 2. Phylogenetic reconstruction of Singapore HIV-1 C2V3C3 sequences with reference strains from CRF01_AE and subtypes B, B′, C, and F1 (outgroup). Numbers at the nodes indicate the percentage of bootstrap values of 2000 replicates.

recent evolution from the founder virus that initiated the Asian CRF01 epidemic. We observed no discordant tree topologies, suggesting that a recombination event with CRF01 had occurred among the subtype B sequences. As discussed above, we observed long branch lengths for S6G316, SG6318, SG6324, SG326, and SG330 CRF01 viruses in the V3 region sequences, suggesting that these strains were more divergent. However, this greater divergence was not reflected as long branches in the corresponding p17 sequences, implying that the two gene regions were evolving in different ways. In fact, the branch length for SG330 was much shorter than that seen for the majority of the Singapore subtype CRF01 p17 sequences, despite the divergent nature of its C2V3C3 sequence.

Amino acid variation within the C2V3C3 region of env in subtype B and CRF01 viruses

Figure 5 shows an alignment of the V3 and flanking regions for the Singapore subtype B and B′ viruses, 1991 subtype B′ sequences from Thailand, and the subtype B reference strains LAI and JRFL. The consensus sequence for subtype B was taken from the Los Alamos database (*35*). Typical of subtype B viruses, there appears to be more variation in the regions immediately

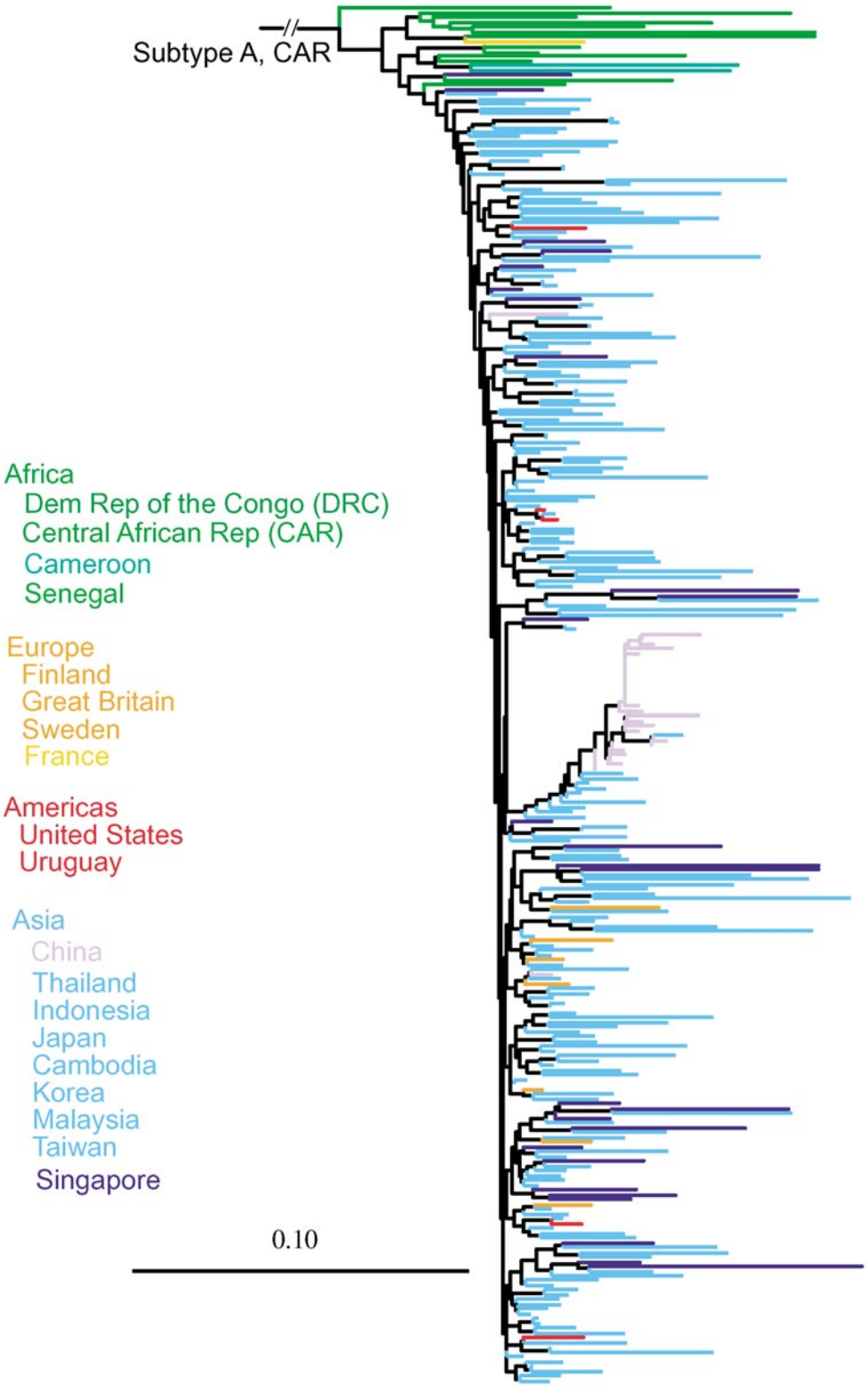

FIG. 3. Phylogenetic reconstruction of CRF01 C2V3C3 sequences from HIV-infected persons living in Africa, Europe, the Americas, and Asia. The outgroups are two subtype A sequences from the Central African Republic. Most Western CRF01 infections are probably of Asian origin, except the French isolate, which clustered with the African CRF01s. There are scattered high bootstrap values throughout the tree between pairs and triplets of sequences, and a 100% bootstrap value for the CRF01 clade compared to the A clade. The high bootstrap values (greater than 70%) for pairs of sequences in most cases support associations from single nations, with the exception of sequences from China and Thailand. This is not surprising given the short sequence length. The specific high bootstrap values in the tree are not marked as they were not particularly informative and the tree branches are very dense. However the trends with nations in the tree are still of interest; the Chinese sequences form a distinctive clade, and the African sequences group near the base of the tree.

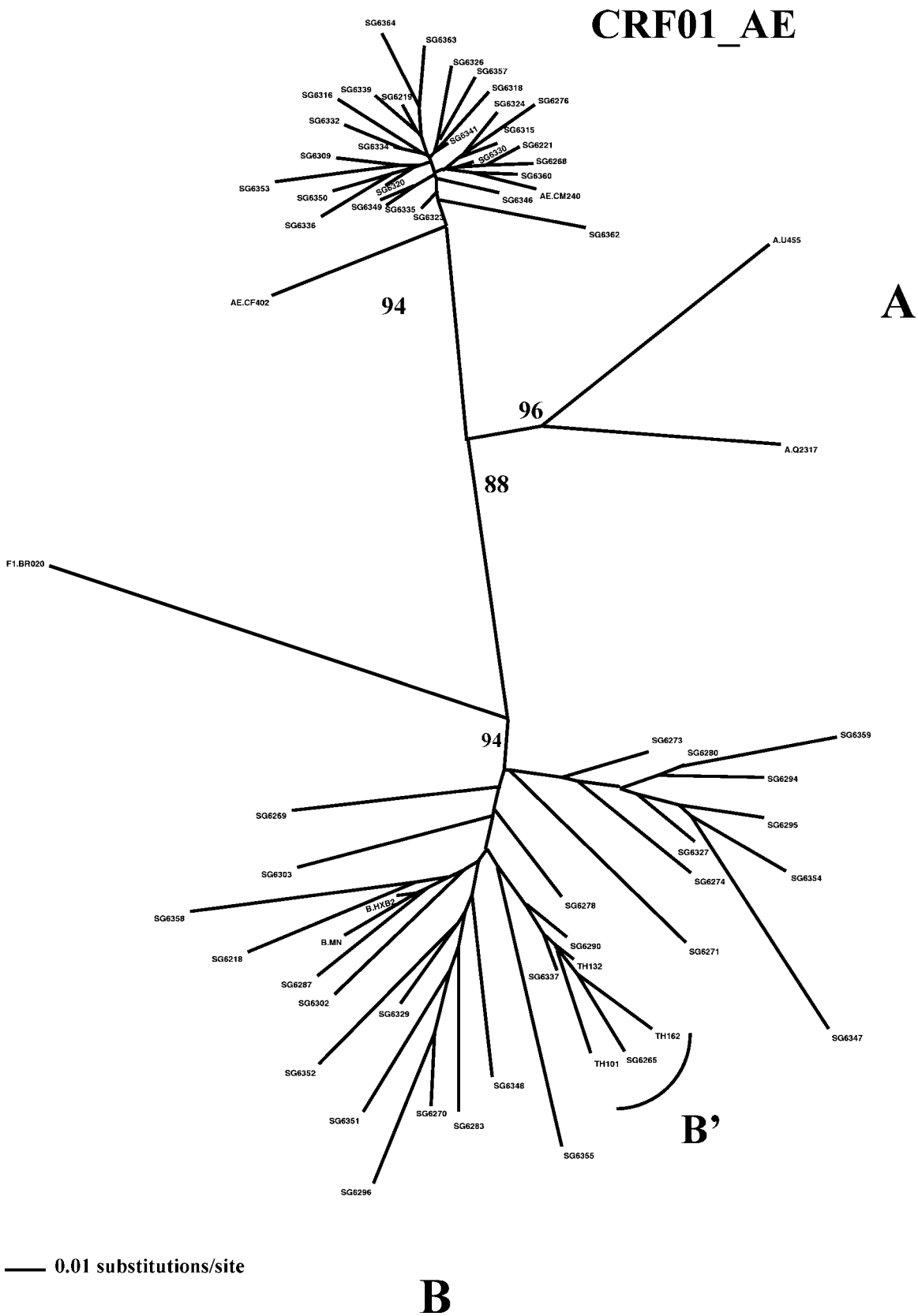

FIG. 4. Phylogenetic reconstruction of Singapore HIV-1 p17 sequences with reference strains from CRF01_AE and subtypes A, B, B′, and F1 (outgroup). Numbers at the nodes indicate the percentage of bootstrap values out of 2000 replicates.

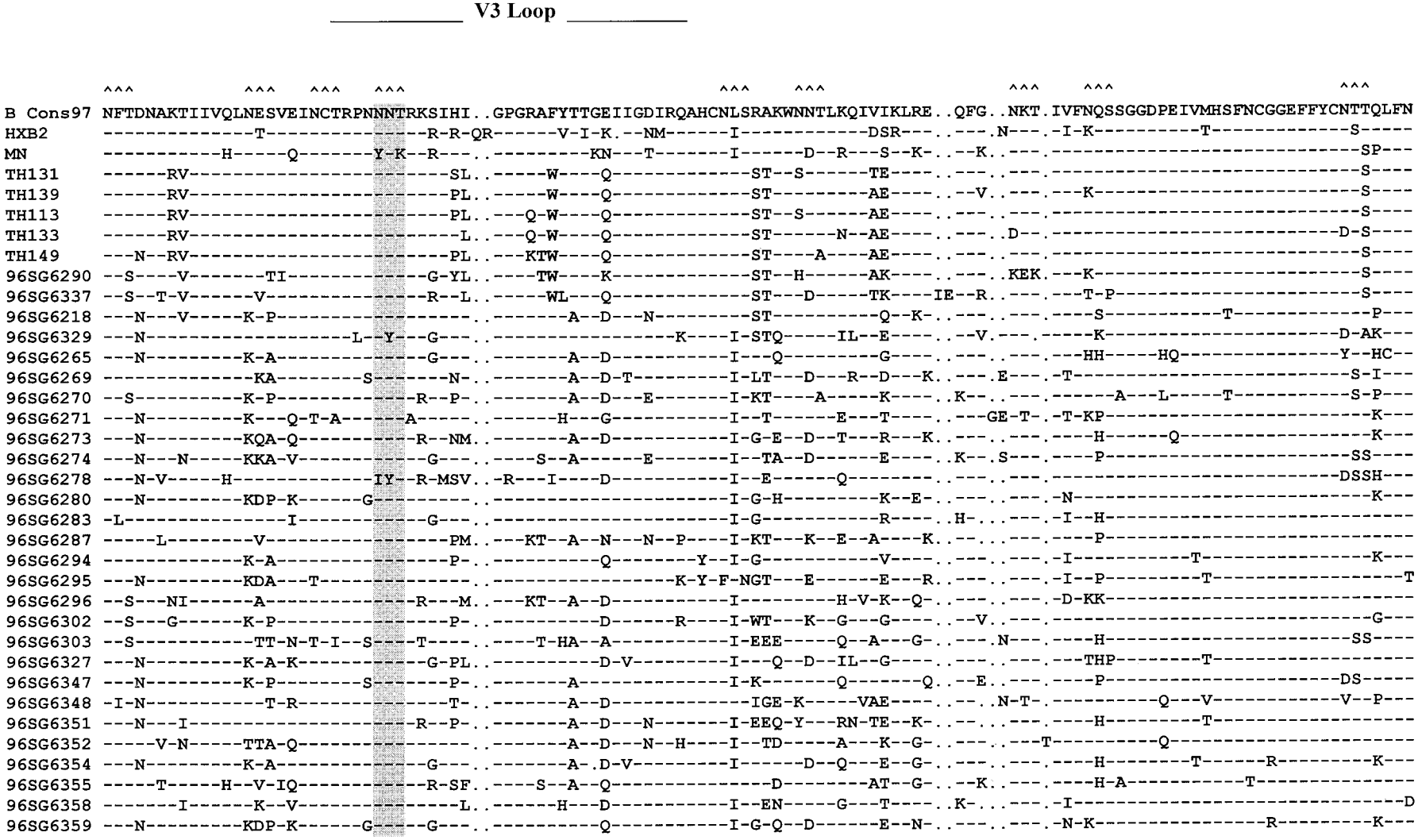

FIG. 5. Amino acid alignment of the Singapore subtype B C2V3C3 sequences. Early Thailand subtype B′, 1996 Singapore, and subtype B reference strains LAI and JRFL are aligned against the 1995 subtype B consensus sequence (*34*). Dashes (- - -) denote identity with the consensus sequence, dots (·) represent insertions/deletions, carets (∧ ∧ ∧) locate potential N-linked glycosylation sites, and positions in gray identify the glycosylation site at position 6 within the V3 loop. All Singapore sequences are identified by the number 96, indicating the year of collection (1996), followed by the country code for Singapore (SG) and by four-digit sequence identifiers.

flanking the third variable region than within the V3 loop itself (*31, 35*), with the greatest number of mutations occurring just downstream of V3. The GPGR motif at the crown of the V3 loop is fairly conserved in subtype B viruses, and this was true for the Singapore B viruses. In previous reports of subtype B′ sequences, 13 (81%) of 16 Thailand B′ viruses collected in 1991 contained a GPGQ motif (*59*), as did 39 (57%) of 69 Thai subtype B′ viruses collected in 1994 (*31*). None of the three Singapore B′ amino acid sequences contained the GPGQ motif, instead, they all had a GPGR motif. For the "typical" subtype B sequences, 25 (86%) of 28 sequences had a GPGR motif, 2 (7%) had a GPGK motif, and 1 (3.6%) had the unusual GRGR motif.

An amino acid alignment of the CRF01 sequences is shown in Fig. 6. Comparison of the amino acid sequences suggested two patterns of V3 loop sequences. While the relative amount of amino acid substitutions, distinct from the CRF01 consensus sequence, was similar in all the flanking region sequences, one subset of V3 loop sequences had only few amino acid changes distinct from the consensus; the other had numerous substitutions. Many of these substitutions resulted in changes to charged amino acids. One example involves the predominant GPGQ motif at the crown of the V3 loop of CRF01 viruses. In every case where there was a mutation from the glutamine (Q), it was to one of the charged amino acids, arginine (R) or histidine (H). In addition, many of the amino acid substitutions in this subset of divergent V3 sequences led to the loss of the glycosylation site typically found at position 6–8 in the V3 loop, and often these also involved changes to charged amino acids.

The role of the glycosylation site at positions 6–8 of the V3 loop on the Singapore CRF01 viruses

Between 18 and 33 N-linked glycosylation sites have been predicted in the gp120 protein of HIV-1 group M viruses on the basis of the presence of glycosylation sequons (*38*); this variation is mainly the result of insertions and deletions in hypervariable regions in the env gene. A potential glycosylation site (glycan) is determined from the sequon asparagine—any amino acid—threonine or serine, N (X) T/S. The subtype B consensus sequence contains eight predicted glycans in the region that we have sequenced (Fig. 5). The second glycosylation site was missing in 19 (68%) of the 28 Singapore subtype B sequences.

It has been reported previously that most sequences of CRF01 viruses lack the fifth, sixth, and eighth glycosylation sites, but that the fifth glycan is replaced just 2 amino acids downstream by a unique alternative glycan (*31, 56, 77*). We also found that 10 (34%) of 29 CRF01 sequences lack the fourth glycosylation site, which begins at the sixth position within the V3 loop; but it is missing in only 1 (3.6%) of the 28 Singapore subtype B sequences. Since glycans have the potential to mask epitopes (*7, 8, 49*), and this glycosylation site is within the V3 loop, we calculated pairwise comparisons for CRF01 C2V3C3 sequences with and without the site (Table 2). We found that the C2V3C3 pairwise distances and *dn/ds* ratios were greater for sequences without the glycosylation site (mean 0.1218, *dn/ds* 1.3337) than for those that contained it (mean 0.0789, *dn/ds* 0.9776). We then calculated pairwise comparisons (Table 2) separately for the V3 loop and flanking regions of CRF01 sequences. The flanking regions were used as internal controls. We found the distribution of diversity scores and *dn/ds* ratios were essentially identical in the flanking regions (mean 0.0868, *dn/ds* 0.9004 with the glycosylation site; mean 0.0959, *dn/ds* 0.9776 without it), independent of the glycosylation site, but extremely different within the V3 loop depending on the presence or absence of the glycosylation site (mean 0.0627, *dn/ds* 1.1001 and mean 0.1841, *dn/ds* 2.1712, respectively).

The influence of the glycosylation site at positions 6–8 of the V3 loop of 60 Asian CRF01 sequences

To determine the influence of this glycosylation site on Asian subtype CRF01 protein sequences, we separated V3 sequences that carried the site from those that did not among a set of 60 CRF01 sequences from Thailand, Taiwan, Vietnam, and Singapore extracted from the Los Alamos database. Of these 60 sequences, 20 (33%) lacked the glycosylation site, similar to the percentage among Singapore CRF01 sequences. Of 126 subtype B sequences from Thailand, Taiwan, and Vietnam, only 6 (5%) lacked the site. We also performed PIMA analysis (*34*) on the 60 Asian subtype CRF01 sequences. PIMA scores quantify the degree of similarity between sequences at the amino acid level. The lower the score of a particular sequence, the more distant it is from the consensus. As can been seen from the histogram in Fig. 7A, the V3 loops for the 60 Asian CRF01 sequences show a wide range of PIMA scores with a complex distribution, suggesting the presence of a diverse population of viruses. When we extracted the sequences that contain the glycosylation site and analyzed them separately, the resulting histogram illustrates that these sequences are highly conserved and the distribution is unimodal (Fig. 7B). The bottom panel, a histogram representing Asian CRF01 sequences lacking the glycosylation site, shows that these sequences are far less conserved and have a broad range of variation (Fig. 7C). If the distribution of nonsynonymous substitutions for the histograms were superimposed on one another, the broad, complex distribution of PIMA scores seen in Fig. 7A would be apparent. We next performed an additional test to better characterize the relationship between the mutational pattern and charge and generated a single PIMA score

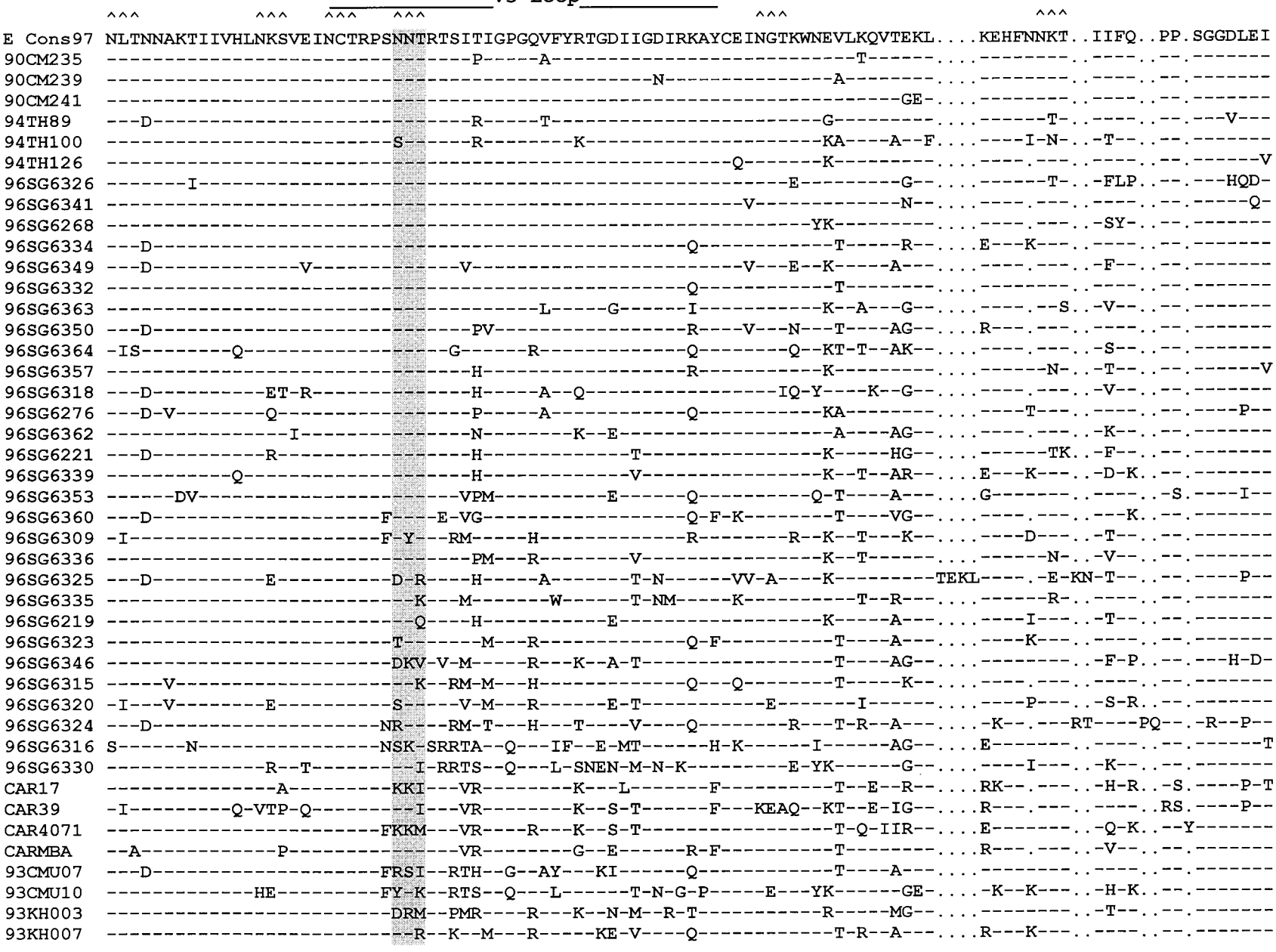
V3 Loop
E Cons97 NLTNNAKTIIVHLNKSVEINCTRPSNNTRTSITIGPGQVFYRTGDIIGDIRKAYCEINGTKWNEVLKQVTEKL....KEHFNNKT..IIFQ..PP.SGGDLEI
90CM235
90CM239
90CM241
94TH89
94TH100
94TH126
96SG6326
96SG6341
96SG6268
96SG6334
96SG6349
96SG6332
96SG6363
96SG6350
96SG6364
96SG6357
96SG6318
96SG6276
96SG6362
96SG6221
96SG6339
96SG6353
96SG6360
96SG6309
96SG6336
96SG6325
96SG6335
96SG6219
96SG6323
96SG6346
96SG6315
96SG6320
96SG6324
96SG6316
96SG6330
CAR17
CAR39
CAR4071
CARMBA
93CMU07
93CMU10
93KH003
93KH007

TABLE 2

Comparisons of Genetic Distance and Synonymous/Nonsynonymous Substitutions within CRF01 and Subtype B Viruses in p17 and C2V3C3

Subtype	Gene region	Distance mean (range)	Nonsynonymous distance (*dn*)	Synonymous distance (*ds*)	*dn*/*ds*
Subtype B	P17	0.0775 (0.0317–0.1263)	0.0558 (0.0000–0.103)	0.1593 (0.0353–0.3452)	0.3503
CRF01	P17	0.0329 (0.0136–0.1924)	0.0352 (0.0103–0.0681)	0.0545 (0.0000–0.1364)	0.6459
Subtype B	C2V3C3	0.1232 (0.0309–0.1924)	0.1199 (0.0353–0.1854)	0.1376 (0.0147–0.323)	0.8712
CRF01	C2V3C3	0.0942 (0.0304–0.2079)	0.0971 (0.0232–0.2252)	0.0841 (0.0139–0.2179)	1.1543
Subtype B	Flanking	0.1306 (0.0359–0.1941)	0.1284 (0.0392–0.1980)	0.1418 (0.0218–0.2899)	0.9055
CRF01	Flanking	0.0892 (0.0309–0.1913)	0.0880 (0.0305–0.1791)	0.0947 (0.0178–0.2902)	0.9292
Subtype B	V3 loop	0.1077 (0.0099–0.2682)	0.1013 (0.0128–0.2446)	0.1348 (0.0000–0.5839)	0.7515
CRF01	V3 loop	0.1073 (0.0000–0.3445)	0.1212 (0.0000–0.3718)	0.0686 (0.0000–0.2625)	1.7667
CRF01	C2V3C3 with glycan[a]	0.0789 (0.0304–0.1475)	0.0785 (0.0232–0.1429)	0.0803 (0.0278–0.2179)	0.9776
	C2V3C3 without glycan[b]	0.1218 (0.0751–0.2079)	0.1291 (0.0653–0.2252)	0.0968 (0.0422–0.1697)	1.3337
	Flanking with glycan[a]	0.0868 (0.0355–0.1913)	0.0850 (0.0334–0.1791)	0.0944 (0.0335–0.2902)	0.9004
	Flanking without glycan[b]	0.0959 (0.0538–0.1532)	0.0958 (0.0563–0.1687)	0.0980 (0.0178–0.2132)	0.9776
	V3 loop with glycan[a]	0.0627 (0.0000–0.1585)	0.0645 (0.000–0.1561)	0.0586 (0.0000–0.2193)	1.1001
	V3 loop without glycan[b]	0.1841 (0.0698–0.3445)	0.2143 (0.0813–0.3718)	0.0987 (0.0000–0.2625)	2.1712

Note. The Asian CRF01 C2V3C3 sequences that were glycosylated at positions 6–8 within the V3 loop had *dn*/*ds* values that were indistinguishable when the region flanking the V3 loop and the V3 loop were compared, (two-sided Wilcoxon rank sum *P*-value 0.7116, $Z = -0.3697$) and the overall selection pressure was indistinguishable in the two regions. On the other hand, in the V3 loop sequences without the glycosylation site, the *dn*/*ds* ratio was roughly twice as high within the loop as it was in the flanking regions (two-sided Wilcoxon rank sum *P*-value 0.0000065, $Z = 4.5091$), close to neutral in the flanking regions and under strong positive selection in the loop. (See table and the figure with the histograms.)

[a] With glycan refers to sequences which contain the glycosylation site at positions 6–8 of the V3 loop.

[b] Without glycan refers to sequences which lack the glycosylation site at positions 6–8 of the V3 loop. Refer to Fig. 5 and 6 for location of the V3 loop.

for each CRF01_AE sequence by comparing it with the CRF01_AE consensus sequence. Again, low PIMA scores indicate multiple nonconservative substitutions relative to the consensus. When we compared this score with the net positive charge of each V3 loop, we found that the variability of the amino acid sequence was strongly correlated with a high net positive charge ($P < 10^{-6}$).

In a further analysis using SNAP to track synonymous and nonsynonymous substitutions for each codon in the alignment, we found that the nonsynonymous substitutions that dominate the V3 loop when the N-linked glycosylation site is absent were focused on either side of the tip of the V3 loop, focused in a region highly susceptible to antibody binding (Fig. 8) (*23, 84*).

DISCUSSION

The first case of HIV infection in the island state of Singapore was reported in May 1985 (*9*). From 1985 to 1990, most HIV infections were in homosexual males, most of whom reported having had sex while in the

FIG. 6. Amino acid alignment of the Singapore CRF01_AE C2V3C3 sequences. Singapore sequences: 1990, 1993, and 1994 sequences from Thailand; and strains from the Central African Republic are aligned against the 1995 subtype E (CRF01_AE) consensus sequence. Dashes (- - -) denote identity with the consensus sequence, dots (•) represent insertions/deletions, carets (∧ ∧ ∧) locate potential N-linked glycosylation sites, and positions in gray identify the glycosylation site at position 6 within the V3 loop. All Singapore sequences are identified by the number 96, indicating the year of collection (1996), followed by the country code for Singapore (SG) and by four-digit sequence identifiers.

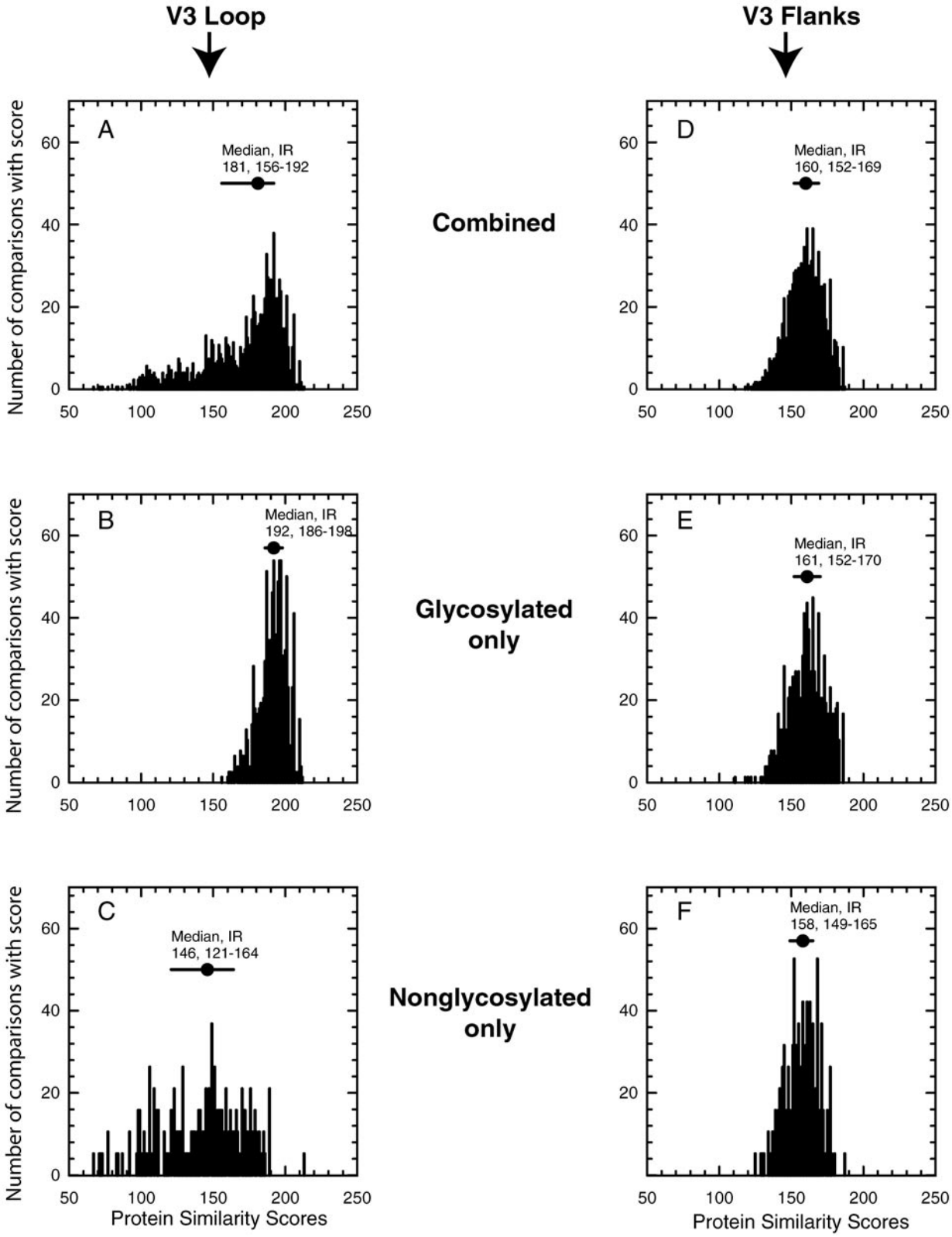

FIG. 7. Pairwise distributions of protein similarity scores for the CRF01V3 loop and the V3 flanking regions show that more divergent V3 loop sequences are associated with the lack of the V3 internal glycosylation site. The sequences included here were the entire 1999 collection of Asian CRF01 sequences from the HIV database that span the V3 loop and flanking regions. The set was restricted to include no more than one sequence per patient. The full distribution and the median and interquartile range (IR) are shown. The sequences were divided into two regions: the V3 loop, typically consisting of 35 amino acid residues, including two cysteine residues that form a disulfide bond to produce the loop; and 35 amino acids flanking the loop (15 amino acids immediately adjacent to the amino-terminal side of the loop and 20 amino acids immediately adjacent to the

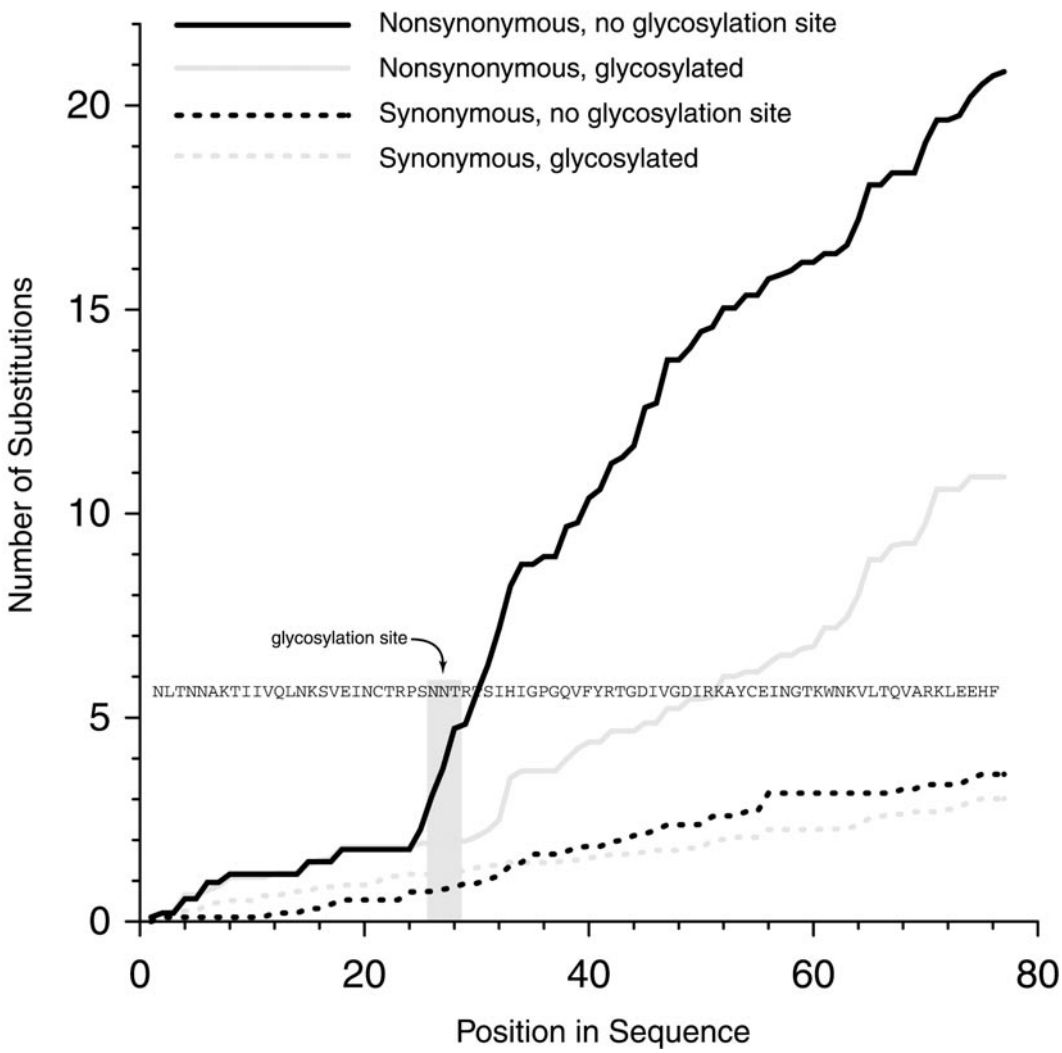

FIG. 8. A comparison of synonymous vs nonsynonymous cumulative behavior, codon by codon, through the V3 region, for the CRF01 sequences with and without the V3 glycosylation site. The slope of each line describes the rate of accumulation of synonymous vs nonsynonymous substitutions as one scans across the V3 region using the subsets of CRF01 V3 loop sequences described in Fig. 6. The average number of either synonymous or nonsynonymous substitution was calculated for each set for each codon, and this number is added to the cumulative score as one moves from left to right. The synonymous substitutions (dashed lines) are accumulating at roughly the same rate for both the CRF01 set that contains the V3 loop glycosylation site and the set that does not contain the glycosylation site. The steep slope of the solid black line near the center of the V3 loop reflects a very high rate of accumulation of nonsynonymous distance in the subset of Asian CRF01 sequences lacking the glycosylation site in the V3 loop, which is indicative of positive selection. The most rapid accumulation of nonsynonymous substitutions begins at the missing glycosylation site, just inside of the amino-terminal side of the V3 loop; for the N-terminal part of the alignments, up to that point, the nonsynonymous substitutions track almost perfectly for the two sets. The high substitution rate for the set lacking the glycosylation site spans the tip of the loop, which continues through the C-terminal cysteine, at which point the two sets begin to track each other once again. (Although the cumulative line for the nonglycosylated set is of course much higher beyond the V3 loop, the rates of nonsynonymous substitution as reflected by the parallel tracking of the lines beyond the V3 loop indicates that the selective pressure is comparable in both sets.).

carboxyl-terminal side of the loop). Each of the two regions was aligned and scored for pairwise similarity by using a modified version of the PIMA software package (*70*) that incorporates protein similarity scores based on amino acid substitution matrices from protein blocks (*25, 33*). (A) The nonnormal distribution with a long tail of low protein similarity indicates that a subset of the Asian E subtype V3 loop sequences are more divergent than the rest of the population. (B) The subset of V3 loop sequences containing a glycosylation site at position six is relatively highly conserved. (C) The subset of V3 loop sequences lacking this glycosylation site is highly divergent. (D, E, F) The flanking regions of the V3 loop serve as an internal control for this analysis. The flanks exhibit roughly the same distribution for the entire set, the glycosylated sequences, and the nonglycosylated sequences, supporting the conclusion that the subset of V3 loop sequences lacking the glycosylation site may be under selective pressure to escape detection by the immune system.

United States, Europe, or Australia, where HIV infections were due almost exclusively to subtype B viruses. In early 1991, the epidemiology of HIV changed, with 74% of all new infections resulting from heterosexual transmissions. By October 1997, 694 Singapore residents were reported to be HIV-infected, with 96% of the infections occurring through casual sex or sex with prostitutes in Singapore and overseas. In this study of the molecular epidemiology of HIV-1 infections in Singapore in 1996, we demonstrated that the subtype B and CRF01 viruses were introduced and spread separately: subtype B through homosexual transmission and CRF01 through heterosexual transmission.

Studying the early dynamics of an evolving regional epidemic involving viruses of more than one subtype provides the opportunity to better understand the factors that may influence the selective transmission of viruses from one subtype within diverse risk groups, whether those factors are behavioral, biological, or simply due to chance. It also presents the opportunity to compare how the viruses in each subtype are evolving within a population having a fairly homogenous racial and genetic background. The early occurrence of HIV infections among homosexual men as early as the mid-1980s suggests that subtype B viruses were introduced into Singapore early in the global HIV pandemic, and the greater genetic diversity that we observed in both the env and the p17 sequences of subtype B viruses suggests that they may have been introduced into Singapore earlier than CRF01. In addition, because homosexual men from Singapore frequently traveled to countries where subtype B viruses were prevalent, multiple introductions of these B strains most likely occurred also. This is consistent with the hypothesis of an early, global founder effect in the rapid dissemination of subtype B viruses by homosexual transmission (*41*).

In contrast, the first recorded date of heterosexual infection in Singapore was in 1991, only 1–2 years after the CRF01 epidemic began in Thailand (*58, 59, 72, 82*). When reporting their travel histories, most heterosexuals in Singapore documented overseas trips to countries within Southeast Asia, where CRF01 viruses account for most heterosexually acquired infections. Whether there are biological differences among subtype B and CRF01 viruses that may explain this apparent segregation remains to be determined (*15, 51–53, 62, 63, 74*). Nonetheless, the epidemiologic evidence seems sufficient to conclude that the earlier and most likely multiple introductions of subtype B viruses occurred through homosexual transmissions and that the later introduction of CRF01_AE viruses occurred through heterosexual transmission.

Comparing the synonymous and nonsynonymous pairwise distances provides a means of assessing evolutionary rates and the role of selection in the evolution of different gene regions within and between subtypes. The balance of positive and negative selective pressures within a host are not equal for all genes or gene regions, hence, they do not evolve at the same rate (*4, 69*). While env is more variable than gag and pol, within each gene there are regions of greater or lesser variability, i.e., the p17 coding region of gag is more variable than p24 (*4*). The similar *dn/ds* ratios for the regions flanking the V3 loop were the same for subtype B and CRF01, demonstrating they are evolving at a constant rate. However, we observed two patterns in the CRF01 V3 loop relative to the subtype B sequences: either highly conserved or very divergent. Hence, a subset of Asian V3 loop sequences are evolving more rapidly than the rest of the V3 region. When evaluating the role of the highly conserved glycosylation site at positions 6–8 within the V3 loop, we observed that 34% of our Singapore CRF01 viruses lacked this site. Glycans have been reported to play an important role in antibody recognition, viral infectivity, cellular host range, and coreceptor usage (*2, 3, 13, 44, 45, 49*); more importantly, they have been reported to mask epitopes (*2, 3*). Therefore, we hypothesized that viruses lacking the glycan would be more susceptible to immune surveillance, i.e., that the V3 loop of these viruses would be under constant selective pressure to mutate and escape immune recognition. Indeed, we found that the V3 loops of sequences lacking the glycosylation site had higher *dn/ds* ratios in the V3 loop than did the sequences containing the glycosylation sites, while the flanking regions from strains with and without the site displayed equal rates of evolution. When we extended our analysis to include 60 CRF01 sequences from Thailand, Taiwan, Vietnam, and Singapore and 126 subtype B sequences from Thailand, Taiwan, and Vietnam, we found a complex distribution of protein similarity scores for all CRF01 sequences, including those from Africa (*14, 34*). We found that V3 protein sequences were highly conserved in viruses that contained the glycosylation site, but those without the site showed a broad range of divergent sequences. It has recently been reported that mutations leading to substitutions with positively charged amino acids in this glycosylation site can also have an influence on changing CRF01 viruses from NSI to SI phenotypes (*33, 68*), further confirming the important role for this glycan in the evolution of the V3 loop of CRF01 viruses. In our study, we found a high positive charge as well as the lack of the glycosylation site to be related to V3 amino acid diversity. A plausible hypothesis that accounts for all of our observations is that the lack of the glycosylation site and the positive charge reflect CXCR4 coreceptor usage, and that the V3 loop in these viruses is thus exposed to more intense immune surveillance. Positive selection caused by immune escape would account for the greater diversity near the tip of the loop that we found in these viruses.

In conclusion, we have shown that subtype B and CRF01 were introduced into Singapore sequentially, sub-

type B through homosexual transmission, and CRF01 through heterosexual transmission. We also determined that a subpopulation of CRF01 viruses were evolving more rapidly in the V3 region and that the loss of a glycosylation site within the V3 loop is strongly associated with this accelerated adaptation. This type of study is important for countries interested in participating in vaccine trials. Thailand was selected as one of the first vaccine trial sites, in part because the Thai CRF01 and subtype B′ epidemic was recently introduced from a single founder virus, and the strains were genetically very similar. However, the existence of this subpopulation of CRF01, whose V3 region is evolving more rapidly, may present an unanticipated challenge to the development of CRF01 vaccines based solely on *gp*120. To gauge the extent of this challenge, researchers will need to determine whether there are biological as well as genotypic differences between persons infected with this more rapidly evolving subpopulation of CRF01 strains and those infected with strains evolving more slowly. These rapidly evolving viruses should also be considered when evaluating breakthrough strains from CRF01 vaccine trials.

MATERIALS AND METHODS

Subjects

Whole blood specimens from 105 HIV-1-infected persons were collected from patients attending the HIV clinic at the Communicable Disease Centre, Singapore. The only criterion for inclusion into the survey was a known HIV-positive infection status. Patient forms containing demographic data, risk information, and latest CD4 count was filled out at the time of blood collection (summarized in Table 1), and the patients' signed informed consent was received. Data and blood samples were coded and sent to the U.S. Centers for Disease Control and Prevention (CDC) as blinded samples.

Viral DNA isolation, amplification, and sequencing

All samples were initially processed in the National Skin Centre, Singapore before being shipped to CDC. Eight milliliters of whole blood was collected from HIV-infected patients into the Vacutainer CPT Cell Preparation Tube with sodium citrate (Becton–Dickinson, Franklin Lakes, NJ) and centrifuged according to the manufacturer's instructions. The blood specimens were processed within 2–3 h of collection. Briefly, the contents of the tubes were mixed by gently inverting the tubes 8 to 10 times and centrifuging them at 1500 to 1800 *g* at room temperature for 20 min. The plasma layer above the band of peripheral blood mononuclear cells (PBMCs) in Ficoll–Hypaque was discarded. The band of PBMCs and Ficoll–Hypaque from above the gel plug was removed to a 15-mL tube, and 1.0 mL of prewarmed red blood cell osmotic buffer, pH 7.4 (prepared by mixing 18 mL 0.17 M NH_4Cl with 2 mL 0.16 M Tris), was added. After the tubes stood for 10 min at room temperature, 9 mL PBS was added and mixed, and the tubes were spun at 250 *g* at room temperature for 15 min. The supernatant was discarded. To the residual cells, 0.5 mL PCR lysis buffer (50 mM KCl, 10 mM Tris–HCl, pH 8.3, 2.5 mM $MgCl_2$, 25 M EDTA, 0.45% NP-40, and 0.45% Tween 20) containing 60 M/mL Proteinase K was added. The cells were lysed at 56°C for 1 h and at 95°C for another 15 min. The extracted DNA specimens were stored at −20°C. Fifty-eight DNA lysates from samples with sufficient cell pellets were used in nested polymerase chain reactions (PCRs) as previously described (*31*). Amplified products were approximately 720 bp starting in the C2 region of the env gene and terminating at the end of gp120, and approximately 474 bp beginning upstream and spanning the length of the *p*17 coding region of the *gag* gene. Primers for the C2V3C3 were MK603 (5′ AGAAAAATGGTGGGTCACAGTCTATTATGGGGTACCT 3′) and CO602 (5′ GCCCATAGTGCTTCCTGCTGCTCCCAAGAACC 3′) for primary PCR, and primers MK650 (5′ ATGTCAGCACAGTACAATGTACAC 3′) and MK601 (5′TTCTCCAATTGTCCCTCATATCTCCTCCTCCA 3′) for nested PCR. The p17 primers were CL1028 (5′ TCTCTAGCAGTGGCGCCCGAACAGGGAC 3′) and CL1033 (5′ TCTATCCCATTCTGCAGCTTCCTCATTGAT 3′) for primary PCR, and primers CL1029 (5′ TTTGACTAGCGGAGGCTAGA 3′) and CL1032 (5′ GGTGATA TGGCCTGATGTACCATTTGCCCCTG 3′) for nested PCR.

PCR products were purified with the Qiagen PCR purification spin kit (Qiagen, Inc., Chatsworth, CA), and the DNA was sequenced according to the *Taq* dye terminator method (Applied Biosystems, Inc., Foster City, CA). Primers used for sequencing approximately 345 bp of the C2V3C3 region of *env* were KR207 (5′ TGTTAAATGGCAGTCTAGC 3′) and MK650. Sequencing primers for sequencing the approximately 396 bp of p17 included CL1029, CL1032, and CL1030 (5′ AGACAGGATCAGAAGAA 3′).

Genetic and phylogenetic analysis

For determining subtype classification and phylogenetic relationships, sequence comparisons for the 28 subtype B and 29 CRF01 strains were performed by using the C2V3C3 region of *env* and the *p*17 coding region of *gag*. The sequences were manually aligned with the sequence editor function of the Genetic Data Environment (GDE) package. Sites where there was a gap in any of the sequences or ambiguous positions in the alignment were excluded from phylogenetic analysis. The Molecular Evolution Genetic Analysis (MEGA) program (*42*) was used to perform pairwise nucleotide comparisons of the sequences, and Synonymous Nonsynonymous Analysis Program (SNAP) (*36*) was used to

estimate synonymous (*ds*) and nonsynonymous (*dn*) distances calculated by the method (*57*). *ds* is the number of observed synonymous substitutions divided by the number of possible synonymous substitutions and *dn* is the number of observed nonsynonymous substitutions divided by the number of possible nonsynonymous substitutions. The nonparametric Wilcoxon rank sum test statistic was used to compare the used distributions of *dn/ds* values in the glycosylated and nonglycosylated V3 loop. All pairwise comparisons of sequences were used to obtain the average *dn/ds* values presented in Table 2, but the non-independence of points and inflated apparent sample size make the sets of all pairwise comparisons inappropriate for statistical comparisons. To address this problem, the average *dn/ds* value was obtained for each sequence compared to all others, and thus, one *dn/ds* value was assigned to each sequence for use in the final comparison.

For phylogenetic tree analysis, separate nucleotide alignments for Singapore *env* C2V3C3 ($n = 58$) and p17 ($n = 56$) sequences were performed with the Se-Al sequence alignment editor v.1.0 (*64*). HIV reference strains from subtypes A, AE, B, B′, C, and F were incorporated into the alignments. Alignments were saved with removal of nucleotide positions containing gaps (336 bp for the C2V3C3 alignment and 363 bp for the p17 alignment). The Modeltest program (*64*) was used with each alignment to test for a statistically justified model of DNA substitution for use in the phylogenetic tree-building program. The model chosen in each case by the program was used as the DNA substitution model in the bootstrap (2000 replicates) (*25*); the neighbor-joining methodology was implemented in PAUP (*79*). The Singapore sequence data were deposited in GenBank under Accession Nos. AF385955–AF386067 and AF417200.

ACKNOWLEDGMENTS

We thank Dr. Thomas M. Folks, NCID, CDC, and Dr. Timothy D. Mastro, NCHSTP, CDC, Atlanta, for helpful discussions and critical review of the manuscript, and Robin Moseley for editing.

REFERENCES

1. Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220,** 868–871.
2. Benjouad, A., Gluckman, J. C., Rochat, H., Montagnier, L., and Bahraoui, E. (1992). Influence of carbohydrate moieties on the immunogenicity of human immunodeficiency virus type 1 recombinant gp160. *J. Virol.* **66,** 2473–2483.
3. Botarelli, P., Houlden, B. A., Haigwood, N. L., Servis, C., Montagna, D., and Abrignani, S. (1991). N-glycosylation of HIV-gp120 may constrain recognition by T lymphocytes. *J. Immunol.* **147,** 3128–3132.
4. Brown, A., and Monaghan, P. (1988). Evolution of the structural proteins of human immunodeficiency virus: Selective constraints on nucleotide substitution. *AIDS Res. Hum. Retroviruses* **4,** 399–407.
5. Bunnell, R. E., Yanpaisarn, S., Kilmarx, P. H., Rhodes, P. H., Limpakarnjanarat, K., Srismith, R., Mastro, T. D., and St Louis, M. E. (1999). HIV-1 seroprevalence among childbearing women in northern Thailand: Monitoring a rapidly evolving epidemic. *AIDS* **13,** 509–515.
6. Carr, J. K., Salminen, M. O., Koch, C., Gotte, D., Artenstein, A. W., Hegerich, P. A., St Louis, D, Burke, D. S., and McCutchan, F. E. (1996). Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* **70,** 5935–5943.
7. Chabot, D. J., Chen, H., Dimitrov, D. S., and Broder, C. C. (2000). N-linked glycosylation of CXCR4 masks coreceptor function for CCR5-dependent human immunodeficiency virus type 1 isolates. *J. Virol.* **74,** 4404–4413.
8. Cheng-Mayer, C., Brown, A., Harouse, J., Luciw, P. A., and Mayer, A. J. (1999). Selection for neutralization resistance of the simian/human immunodeficiency virus SHIVSF33A variant in vivo by virtue of sequence changes in the extracellular envelope glycoprotein that modify N-linked glycosylation. *J. Virol.* **73,** 5294–5300.
9. Kai, C. S. (1985). HIV infection in Singapore. *Epidemiol. News Bull.* **23,** 67–72.
10. Choopanya, K., Vanichseni, S., Des Jarlais, D. C., Plangsringarm, K., Sonchai, W., Carballo, M., Friedmann, P., and Friedman, R. S. (1991). Risk factors and HIV seropositivity among injecting drug users in Bangkok. *AIDS* **5,** 1509–1513.
11. Corbet, S., Muller-Trutwin, M. C., Versmisse, P., Delarue, S., Ayouba, A., Lewis, J., Brunak, S., Martin, P., Brun-Vezinet, F., Simon, F., Barre-Sinoussi, F., and Mauclere, P. (2000). *env* sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. *J. Virol.* **74,** 529–534.
12. De Jong, J. J., De Ronde, A., Keulen, W., Tersmette, M., and Goudsmit, J. (1992). Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: Analysis by single amino acid substitution. *J. Virol.* **66,** 6777–6780.
13. Dedera, D. A., Gu, R. L., and Ratner, L. (1992). Role of asparagine-linked glycosylation in human immunodeficiency virus type 1 transmembrane envelope function. *Virology* **187,** 377–382.
14. Dighe, P. K., Korber, B., and Foley, B. T. (1997). Global Variation in the HIV-1 V3 Region, *In* "Human Retroviruses and AIDS: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences" (B. Korber, B. Hahn, B. Foley, J. W. Mellors, T. Leitner, G. Myers, F. McCutchan, and C. L. Kuiken, Eds.), Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM. [http://hiv-web.lanl.gov]
15. Dittmar, M. T., Simmons, G., Hibbitts, S., O'Hare, M., Louisirirotchanakul, S., Beddows, S., Weber, J., Clapham, P. R., and Weiss, R. A. (1997). Langerhans cell tropism of human immunodeficiency virus type 1 subtype A through F isolates derived from different transmission groups. *J. Virol.* **71,** 8008–8013.
16. Dumitrescu, O., Kalish, M. L., Kliks, S. C., Bandea, C. I., and Levy, J. A. (1994). Characterization of human immunodeficiency virus type 1 isolates from children in Romania: Identification of a new envelope subtype. *J. Inf. Dis.* **169,** 281–288.
17. European Commission, and the Joint United Nations Programme on HIV/AIDS AIDS. (1997). **11**, S17–S36.
18. Gallo, R. C., Sarin, P. S., Gelmann, E. P., Robert-Guroffan, M., Richardson, D. E., Kalyanaraman, V. S., Mann, D., Sidhu, D. G., Stahl, R. E., Zolla-Pazner, S., Leibowitch, J., and Popovic, M. (1983). Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **220,** 865–867.
19. Gao, F., Robertson, D. L., Morrison, S. G., Hui, H., Craig, S., Decker,

J., Fultz, P. N., Girard, M., Shaw, G. M., Hahn, B. H., and Sharp, P. M. (1994). Genetic variation of HIV type 1 in four World Health Organization-sponsored vaccine evaluation sites: Generation of functional envelope (glycoprotein 160) clones representative of sequence subtypes A, B, C, and E. WHO Network for HIV Isolation and Characterization. *AIDS Res. Hum. Retroviruses* **10,** 1359–1368.

20. Gao, F., Morrison, S. G., Robertson, D. L., Thornton, C. L., Craig, S., Karlsson, G., Sodroski, J., Morgado, M., Galvao-Castro, B., and von Briesen, H. (1996a). Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID Networks for HIV Isolation and Characterization. *J. Virol.* **70,** 1651–1667.
21. Gao, F., Robertson, D. L., Morrison, S. G., Hui, H., Craig, S., Decker, J., Fultz, P. N., Girard, M., Shaw, G. M., Hahn, B. H., and Sharp, P. M. (1996b). The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol.* **70,** 7013–7029.
22. Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397,** 436–441.
23. Gorny, M. K., Xu, J. Y., Gianakakos, V., Karwowska, S., Williams, C., Sheppard, H. W., Hanson, C. V., and Zolla-Pazner, S. (1991). Production of site-selected neutralizing human monoclonal antibodies against the third variable domain of the human immunodeficiency virus type 1 envelope glycoprotein. *Proc. Natl. Acad. Sci. USA* **88,** 3238–3242.
24. Hahn, B. H., Shaw, G. M., De Cock, K. M., and Sharp, P. M. (2000). AIDS as a zoonosis: Scientific and public health implications. *Science* **287,** 607–614.
25. Hedges, S. B. (1992). The number of replicates needed for accurate estimation of bootstrap *P* value in phylogenetic studies. *Mol. Biol. Evol.* **9,** 366–369.
26. Henikoff, S., and Henikoff, J. G. (1993). Performance evaluation of amino acid substitution matrices. *Proteins* **17,** 49–61.
27. Holmes, E. C., Zhang, L. Q., Simmonds, P., Ludlam, C. A., and Brown, A. J. (1992). Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc. Natl. Acad. Sci. USA* **89,** 4835–4839.
28. Hu, D. J., Dondero, T. J., Rayfield, M. A., George, J. R., Schochetman, G., Jaffe, H. W., Luo, C. C., Kalish, M. L., Weniger, B. G., Pau, C. P., Schable, C. A., and Curran, J. W. (1996). The emerging genetic diversity of HIV. The importance of global surveillance for diagnostics, research, and prevention. *JAMA* **275,** 210–216.
29. Janssens, W., Heyndrickx, L., Fransen, K., Motte, J., Peeters, M., Nkengasong, J. N., Ndumbe, P. M., Delaporte, E., Perret, J. L., and Atende, C. (1994). Genetic and phylogenetic analysis of env subtypes G and H in central Africa. *AIDS Res. Hum. Retroviruses* **10,** 877–879.
30. Kalish, M. L., Luo, C. C., Weniger, B. G., Limpakarnjanarat, K., Young, N., Ou, C. Y., and Schochetman, G. (1994). Early HIV type 1 strains in Thailand were not responsible for the current epidemic. *AIDS Res. Hum. Retroviruses* **10,** 1573–1575.
31. Kalish, M. L., Baldwin, A., Raktham, S., Wasi, C., Luo, C. C., Schochetman, G., Mastro, T. D., Young, N. L., Vanichseni, S., Rubsamen-Waigmann, H., von Briesen, H., Mullins, J. I., Delwart, E., Herring, B., Esparza, J., Heyward, W. L., and Osmanov, S. (1995). The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: Implications for HIV vaccine trials. *AIDS* **9,** 851–857.
32. Kasper, P., Kaiser, R., Oldenburg, J., Brackmann, H. H., Matz, B., and Schneweis, K. E. (1994). Parallel evolution in the V3 region of HIV type 1 after infection of hemophiliacs from a homogeneous source. *AIDS Res. Hum. Retroviruses* **10,** 1669–1678.
33. Kato, K., Sato, H., and Takebe, Y. (1999). Role of naturally occurring basic amino acid substitutions in the human immunodeficiency virus type 1 subtype E envelope V3 loop on viral coreceptor usage and cell tropism. *J. Virol.* **73,** 5520–5526.
34. Korber, B. T., MacInnes, K, Smith, R. F., and Myers, G. (1994). Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1. *J. Virol.* **68,** 6730–6744.
35. Korber, B., Hahn, B., Foley, B., Mellors, J. W., Leitner, T., Myers, G., McCutchan, F., and Kuiken, C. L. (1997). Human Retroviruses and AIDS: A compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM. [http://hiv-web.lanl.gov]
36. Korber, B. (2000). HIV signature, and sequence variation analysis *In* "Computational Analysis of HIV Molecular Sequences" (A. G. Rodrigo and G. H. Learn, Eds.), Kluwer Academic Publishers, Dordrecht, Netherlands. [Interactive SNAP program available at http:hiv-web.lanl.gov]
37. Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., and Wolinsky, S. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science* **288,** 1789–1796.
38. Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., and Detours, V. (2001). Evolutionary, and immunological implications of contemporary HIV-1 variation. *In* "The Changing Face of HIV and AIDS" (R. Weiss, M. Adler, and S. Rowland-Jones, Eds.), British Medical Bulletin, Oxford Univ. Press, London.
39. Kuiken, C. L., Lukashov, V. V., Baan, E., Dekker, J., Leunissen, J. A., and Goudsmit, J. (1996). Evidence for limited within-person evolution of the V3 domain of the HIV-1 envelope in the Amsterdam population. *AIDS* **10,** 31–37.
40. Kuiken, C. L., Foley, B., Hahn, B., Korber, B., McCutchan, F., Marx, P. A., Mellors, J. W., Mullins, J. I., Sodroski, J., and Wolinksy, S. (1999). Human retroviruses and AIDS: A compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM. [http://hiv-web.lanl.gov]
41. Kuiken, C., Thakallapalli, R., Esklid, A., and De Ronde, A. (2000). Genetic analysis reveals epidemiologic patterns in the spread of human immunodeficiency virus. *Am. J. Epidemiol.* **152,** 814–822.
42. Kumar, S., Tamura, K., Jakobsen, I., and Nei, M. (1993). MEGA: Molecular evolutionary genetics analysis, version 1.01. The Pennsylvania State University, University Park, PA.
43. Labrosse, B., Treboute, C., Brelot, A., and Alizon, M. (2001). Immunodeficiency virus type 1 gp120 for interaction with the CXCR4 receptor. *J. Virol.* **75,** 5457–5464.
44. Lee, W. R., Syu, W. J., Du, W. J., Matsuda, M., Tan, S., Wolf, A., Essex, M., and Lee, T. H. (1992). Nonrandom distribution of gp120 N-linked glycosylation sites important for infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* **89,** 2213–2217.
45. Li, Y., Luo, L, Rasool, N., and Kang, C. Y. (1993). Glycosylation is necessary for the correct folding of human immunodeficiency virus gp120 in CD4 binding. *J. Virol.* **67,** 584–588.
46. Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W., and Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **73,** 152–160.
47. Louwagie, J., McCutchan, F. E., Peeters, M., Brennan, T. P., Sanders-Buell, E., Eddy, G. A., van der Groen, G., Fransen, K., Gershy-Damet, G. M., and Deleys, R. (1993). Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS* **7,** 769–780.
48. Lukashov, V. V., and Goudsmit, J. (1997). Evolution of the human immunodeficiency virus type 1 subtype-specific V3 domain is

confined to a sequence space with a fixed distance to the subtype consensus. *J. Virol.* **71,** 6332–6338.

49. Malenbaum, S. E., Yang, D., Cavacini, L., Posner, M., Robinson, J., and Cheng-Mayer, C. (2000). The N-terminal V3 loop glycan modulates the interaction of clade A and B human immunodeficiency virus type 1 envelopes with CD4 and chemokine receptors. *J. Virol.* **74,** 11008–11016.
50. Mason, C. J., Kitsiripornchai, S., Markowitz, L. E., Chanbancherd, P., Supapongse, T., Jugsudee, A., Sirisopana, N., Chuenchitra, C., Torugsa, K., VanCott, T. C., Michael, R. A., and Nitayaphan, S. (1998). Nationwide surveillance of HIV-1 prevalence and subtype in young Thai men. *J. Acquir. Immune Defic. Syndr.* **19,** 165–173.
51. Mastro, T. D., Satten, G. A., Nopkesorn, T., Sangkharomya, S., and Longini, I. M., Jr. (1994). Probability of female-to-male transmission of HIV-1 in Thailand. *Lancet* **343,** 204–207.
52. Mastro, T. D., and de Vincenzi, I. (1996). Probabilities of sexual HIV-1 transmission. *AIDS* **10,** S75–SS82.
53. Mastro T. D., Kunanusont, C., Dondero, T. J., and Wasi, C. (1997). Why do HIV-1 subtypes segregate among persons with different risk behaviors in South Africa and Thailand? *AIDS* **11**, 113–116.
54. McCutchan, F. E., Hegerich, P. A., Brennan, T. P., Phanuphak, P., Singharaj, P., Jugsudee, A., Berman, P. W., Gary, A. M., Fowler, A. K., and Burke, D. S. (1992). Genetic variants of HIV-1 in Thailand. *AIDS Res. Hum. Retroviruses* **8,** 1887–1895.
55. McCutchan, F. E., Salminen, M. O., Carr, J. K., and Burke, D. S. (1996). HIV-1 genetic diversity. *AIDS* **10,** S13–S20.
56. Murphy, E., Korber, B., Georges-Courbot, M. C., You, B., Pinter, A., Cook, D., Kieny, M. P., Georges, A., Mathiot, C., and Barre-Sinoussi, F. (1993). Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the Central African Republic. *AIDS Res. Hum. Retroviruses* **9,** 997–1006.
57. Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3,** 418–426.
58. Ou, C. Y., Takebe, Y., Luo, C. C., Kalish, M., Auwanit, W., Bandea, C., de la Torre, N., Moore, J. L., Schochetman, G., and Yamazaki, S. (1992). Wide distribution of two subtypes of HIV-1 in Thailand. *AIDS Res. Hum. Retroviruses* **8,** 1471–1472.
59. Ou, C. Y., Takebe, Y., Weniger, B. G., Luo, C. C., Kalish, M. L., Auwanit, W., Yamazaki, S., Gayle, H. D., Young, N. L., and Schochetman, G. (1993). Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. *Lancet* **341,** 1171–1174.
60. Peeters, M., Honore, C., Huet, T., Bedjabaga, L., Ossari, S., Bussi, P., Cooper, R. W., and Delaporte, E. (1989). Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS* **3,** 625–630.
61. Peeters, M., Fransen, K., Delaporte, E., Van den Haesevelde, M., Gershy-Damet, G. M., Kestens, L., van der Groen, G., and Piot, P. (1992). Isolation and characterization of a new chimpanzee lentivirus (simian immunodeficiency virus isolate cpz-ant) from a wild-captured chimpanzee. *AIDS* **6,** 447–451.
62. Pope, M., Frankel, S. S., Mascola, J. R., Trkola, A., Isdell, F., Birx, D. L., Burke, D. S., Ho, D. D., and Moore, J. P. (1997). Human immunodeficiency virus type 1 strains of subtypes B and E replicate in cutaneous dendritic cell-T-cell mixtures without displaying subtype-specific tropism. *J. Virol.* **71,** 8001–8007.
63. Pope, M., HO, D. D., Moore, J. P., Weber, J., Dittmar, M. T., and Weiss, R. A. (1997). Different subtypes of HIV-1 and cutaneous dendritic cells [letter]. *Science* **278,** 786–788.
64. Posada, D., and Crandall, K. A. (1998). Modeltest: Testing the model of DNA substitution. *Bioinformatics* **14,** 817–818.
65. Rambaut, A. (1996). Se-Al: Sequence Alignment Ed., ver.1.0. Oxford, U.K. [Hyperlink: http://evolve.zoo.ox.ac.uk/Se-Al/]
66. Robertson, D. L., Anderson, Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Kalish, M. L., Salminen, M., Sharp, P. M., Wolinsky, S., and Korber, B. (1999). HIV-1 Nomenclature Proposal, *In* "Human Retroviruses and AIDS: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences" (B. Korber, B. Hahn, B. Foley, J. W. Mellors, T. Leitner, G. Myers, F. McCutchan, and C. L. Kuiken, C. L. (Ed.), pp. 492–505. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM [http://hiv-web.lanl.gov]
67. Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F. E., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S., and Korber, B. (2000). HIV-1 nomenclature proposal. *Science* **288,** 55–56.
68. Sato, H., Shiino, T., Kodaka, N., Taniguchi, K., Tomita, Y., Kato, K., Miyakuni, N., and Takebe, Y. (1999). Evolution and biological characterization of human immunodeficiency virus type 1 subtype E gp120 V3 sequences following horizontal and vertical virus transmission in a single family. *J. Virol.* **73,** 3551–3559.
69. Seibert, S. A.; Howell, C. Y.; Hughes, M. K.; Hughes, A. L. (1995). Natural selection on the gag, pol, and env genes of human immunodeficiency virus 1 (HIV-1). *Mol. Biol. Evol.* **12,** 803–813.
70. Shiino, T., Kato, K., Kodaka, N., Miyakuni, T., Takebe, Y., and Sato, H. (2000). A group of V3 sequences from human immunodeficiency virus type 1 subtype E non-syncytium-inducing, CCR5-using variants are resistant to positive selection pressure. *J. Virol.* **74,** 1069–1078.
71. Sirisopana, N., Torugsa, K., Mason, C. J., Markowitz, L. E., Jugsudee, A., Supapongse, T., Chuenchitra, C., Michael, R. A., Burke, D. S., Singharaj, P., Johnson, A. E., McNeil, J. G., McCutchan, F. E., and Carr, J. K. (1996). Correlates of HIV-1 seropositivity among young men in Thailand. *J. Acquir. Immune Defic. Syndr.* **11,** 492–498.
72. Smith, D. G. (1990). Thailand: AIDS crisis looms. *Lancet* **335,** 781–782.
73. Smith, R. F., and Smith, T. F. (1992). Pattern-induced multi-sequence alignment (PIMA) algorithm employing secondary structure-dependent gap penalties for use in comparative protein modeling. *Protein Eng.* **5,** 35–41.
74. Soto-Ramirez, L. E., Renjifo, B., McLane, M. F., Marlink, R., O'Hara, C., Sutthent, S., Wasi, C., Vithayasai, P., Vithayasai, V., Apichartpiyakul, C., Auewarakul, P., Cruz, V., Chui, D. S., Osathanondh, R., Mayer, K., Lee, T. H., and Essex, M. (1996). HIV-1 Langerhans' cell tropism associated with heterosexual transmission of HIV. *Science* **271,** 1291–1293.
75. Speck, R. F., Wehrly, K., Platt, E. J., Atchison, R. E., Charo, I. F., Kabat, D., Chesebro, B., and Goldsmith, M. A. (1997). Selective employment of chemokine receptors as human immunodeficiency virus type 1 coreceptors determined by individual amino acids within the envelope V3 loop. *J. Virol.* **71,** 7136–7139.
76. Strunnikova, N., Ray, S. C., Livingston, R. A., Rubalcaba, E., and Viscidi, R. P. (1995). Convergent evolution within the V3 loop domain of human immunodeficiency virus type 1 in association with disease progression. *J. Virol.* **69,** 7548–7558.
77. Subbarao, S., Limpakarnjanarat, K., Mastro, T. D., Bhumisawasdi, J., Warachit, P., Jayavasu, C., Young, N. L., Luo, C. C., Shaffer, N., Kalish, M. L., and Schochetman, G. (1998). HIV type 1 in Thailand, 1994–1995: Persistence of two subtypes with low genetic diversity. *AIDS Res. Hum. Retroviruses* **14,** 319–327.
78. Subbarao, S., Vanichseni, S., Hu, D. J., Kitayaporn, D., Choopanya, K., Raktham, S., Young, N. L., Wasi, C., Sutthent, R., Luo, C. C., Ramos, A., and Mastro, T. D. (2000). Genetic characterization of HIV-1 subtype E and B strains from a prospective cohort of injecting drugs users in Bangkok, Thailand. *AIDS Res. Hum. Retroviruses* **16,** 699–707.
79. Swofford, D. L. (2001). PAUP, phylogenetic analysis using parsimony, ver.4.0b8. Sinauer & Assoc., Sunderland, MA.

80. Triques, K., Bourgeois, A., Saragosti, S., Vidal, N., Mpoudi-Ngole, E., Nzilambi, N., Apetrei, C., Ekwalanga, M., Delaporte, E., and Peeters, M. (1999). High diversity of HIV-1 subtype F strains in Central Africa. *Virology* **259,** 99–109.

81. Vanichseni, S., Kitayporn, D., Mastro, T. D., Mock, P. A., Raktham, S., Des Jarlais, D. C., Sujarita, S., Srisuwanvilai, L., Young, N. L., Wasi, C., Subbarao, S., Heyward, W. L., Esparza, J., and Choopanya, K. (2001). Continued high HIV-1 incidence in a vaccine trial preparatory cohort of injection drug users in Bangkok, Thailand. *AIDS* **15,** 397–405.

82. Weniger, B. G., Limpakarnjanarat, K., Ungchusak, K., Thanprasertsuk, S., Choopanya, K., Vanichseni, S., Uneklabh, T., Thongcharoen, P., and Wasi, C. (1991). The epidemiology of HIV infection and AIDS in Thailand. *AIDS* **5,** S71–S85.

83. Yu, X. F., Wang, Z., Beyrer, C., Celentano, D. D., Khamboonruang, C., and Nelson, K. (1997). Diversity of full-length subtype E HIV type 1 env sequences in early seroconvertors from northern Thailand. *AIDS Res. Hum. Retroviruses* **13,** 1439–1441.

84. Zolla-Pazner, S., Gorny, M. K., Nyambi, P. N., VanCott, T. C., and Nadas, A. (1999). Immunotyping of human immunodeficiency virus type 1 (HIV): An approach to immunologic classification of HIV. *Virology* **73,** 4042–4051.